Effective Altruism and Collective Obligations

ALEXANDER DIETZ

University of Southern California

Effective altruism (EA) is a movement devoted to the idea of doing good in the most effective way possible. EA has been the target of a number of critiques. In this article, I focus on one prominent critique: that EA fails to acknowledge the importance of institutional change. One version of this critique claims that EA relies on an overly individualistic approach to ethics. Defenders of EA have objected that this charge either fails to identify a problem with EA's core idea that each of us should do the most good we can, or makes unreasonable claims about what we should do. However, I argue that we can understand the critique in a way that is well motivated, and that can avoid these objections.

I. INTRODUCTION

Effective altruism (EA) is a young social movement devoted to the idea of doing good in the most effective way possible.[1] According to William MacAskill, one of its founders, EA 'is about asking, "How can I make the biggest difference I can?" and using evidence and careful reasoning to try to find an answer'.[2]

This movement has been the target of a number of critiques.[3] In this article, I will focus on one prominent critique: the charge that EA fails to acknowledge the importance of institutional change. According to this *institutional critique*, effective altruists (EAs) focus only on how they can do the most good *within* existing political and economic institutions, and therefore neglect the good that could be done by reforming these institutions.[4]

[1] For introductions to effective altruism, see William MacAskill, *Doing Good Better* (London, 2016); Peter Singer, *The Most Good You Can Do* (London, 2015).

[2] MacAskill, *Doing Good Better*, pp. 14–15.

[3] For overviews, see Iason Gabriel, 'Effective Altruism and its Critics', *Journal of Applied Philosophy* 34 (2017), pp. 457–73; Jeff McMahan, 'Philosophical Critiques of Effective Altruism', *The Philosophers' Magazine* 73 (2016), pp. 92–9.

[4] See Amia Srinivasan, 'Stop the Robot Apocalypse', *London Review of Books* 37 (2015), pp. 3–6; Lisa Herzog, 'Can "Effective Altruism" Really Change the World?', *openDemocracy*, <http://www.opendemocracy.net/transformation/lisa-herzog/can-effective-altruism-really-change-world> (2016); Gabriel, 'Effective Altruism and its Critics'. For responses, see Robert Wiblin, 'Effective Altruists Love Systemic Change', *80,000 Hours*, <http://80000hours.org/2015/07/effective-altruists-love-systemic-change/> (2015);

Utilitas: page 1 of 10
doi:10.1017/S0953820818000158

Brian Berkey has recently argued that the institutional critique either fails to identify a problem with EA's core idea that each of us should do the most good we can, or makes unreasonable claims about what we should do.[5] In this article, I argue that one version of this critique does in fact represent an effective challenge to EA. On this version of the critique, EA relies on an overly individualistic approach to ethics, neglecting the importance of our collective obligations, or obligations that are possessed by groups of people as such, rather than by individuals.[6] I argue that EAs have good reasons to take collective obligations seriously. And, I argue, critics do not need to claim that these collective obligations imply that EA is incorrect in its claims about our individual obligations. Instead, I suggest, the problem is not that EA is incorrect, but simply that it is incomplete. EAs should care not only about how each of us can do the most good as individuals, but also about how we can do the most good together.

II. THE INSTITUTIONAL CRITIQUE

Again, the institutional critique claims that EAs neglect the importance of reforming political and economic institutions. EAs often claim that one of the most effective ways to do good is to help people living in extreme poverty. According to Iason Gabriel, many of the changes we could make that would most help these people are 'institutional or systemic' in nature.[7] For example, Gabriel suggests that trade reform and better financial regulation could lift millions of people out of poverty; others suggest much more radical examples, such as anti-capitalist revolution.[8] But despite their good intentions, critics charge, EAs do not devote sufficient effort to bringing about such changes, and may even make them less likely.[9]

One version of this critique appeals to the idea of collective obligations, or obligations that are possessed by groups of people as such. In particular, some critics suggest that certain groups, such as

McMahan, 'Philosophical Critiques'; Brian Berkey, 'The Institutional Critique of Effective Altruism', *Utilitas*, 30 (2018), pp. 143–71.

[5] Berkey, 'Institutional Critique'.

[6] See Frank Jackson, 'Group Morality', *Metaphysics and Morality*, ed. Philip Pettit, Richard Sylvan and Jean Norman (Oxford, 1987), pp. 91–110; Derek Parfit, 'What We Together Do' (ms.); Alexander Dietz, 'What We Together Ought to Do', *Ethics* 126 (2016), pp. 955–82.

[7] Gabriel, 'Effective Altruism and its Critics', p. 468.

[8] See Matthew Snow, 'Against Charity', *Jacobin*, <http://www.jacobinmag.com/2015/08/peter-singer-charity-effective-altruism/> (2015); and Srinivasan, 'Robot Apocalypse'.

[9] Gabriel, 'Effective Altruism and its Critics', p. 468.

the global rich, have an obligation to enact institutional reforms.[10] And if you are a member of a group that has an obligation to enact institutional reforms, we might think, then you individually have an obligation, or at least a strong reason, to do your part in helping to bring about those reforms.[11] But EA, critics charge, is 'profoundly individualistic', and therefore ignores such considerations.[12]

Berkey argues, however, that critics of EA cannot successfully defend these claims. We can summarize Berkey's argument as offering a dilemma to proponents of the critique.

On the one hand, as Gabriel notes, institutional reforms, such as reforms to central trade or financial policies, have very large-scale effects. So if my efforts would be instrumental in bringing about some good institutional reform, this is a good candidate for being the most effective way that I can do good. But in that case, Berkey argues, the idea that we should promote institutional reform is consistent with EA.[13] It may be that EAs in fact tend to ignore these opportunities, but this would not indicate anything wrong with the core commitment of EA.

On the other hand, Berkey argues, if my efforts will not be instrumental in bringing about the institutional reform – for example, because there will not be enough support to bring about the change despite my efforts – then it seems unreasonable to claim that I ought to devote my efforts there rather than elsewhere. It seems unreasonable, that is, to claim that instead of meeting the urgent needs of people or animals, I ought to spend my limited resources in a way that will be futile, because too few others will support the reform in question.[14]

In short, EAs can indeed recommend that individuals devote themselves to institutional reforms, when doing so would be instrumental in bringing about these reforms. And when this is not the case, it is implausible that individuals should devote themselves to institutional reforms rather than to causes where their efforts would do more good.

[10] Srinivasan, 'Robot Apocalypse'; Herzog, 'Can "Effective Altruism" Really Change the World?'.

[11] Herzog, 'Can "Effective Altruism" Really Change the World?'; Berkey, 'Institutional Critique', pp. 12–13.

[12] Srinivasan, 'Robot Apocalypse'.

[13] Berkey, 'Institutional Critique', pp. 153–4. See also Peter Singer, 'The Logic of Effective Altruism', *Boston Review*, <http://www.bostonreview.net/forum/peter-singer-logic-effective-altruism> (2015); Wiblin, 'Effective Altruists Love Systemic Change'.

[14] Berkey, 'Institutional Critique', p. 158.

III. WHY EFFECTIVE ALTRUISTS SHOULD TAKE COLLECTIVE OBLIGATIONS SERIOUSLY

I think the appeal to collective obligations has more going for it than Berkey gives it credit for. I will first motivate this appeal by showing how the issue of institutional reform can be seen as a symptom of a more basic underlying problem. I will then suggest how we can avoid this problem without running into Berkey's objections.

On my proposal, the problem identified by the institutional critique is a special case of a more basic problem first identified in the philosophical literature by Allan Gibbard, and discussed at length by Donald Regan.[15] While Gibbard and Regan present this problem as a challenge for act-utilitarianism, their point straightforwardly applies to EA as well. The problem is that even when we each individually produce the best outcome we can, given what other people are doing, we may together be producing an outcome that is worse than one we could have produced.

Gibbard and Regan illustrate this problem by discussing cases with the following structure. Suppose that you and I can each choose between two actions, A and B. Depending on what we do, we will produce one of four possible outcomes:

		You	
		do A	do B
I	do A	Second-best	Bad
	do B	Bad	Best

Suppose that we both do A. In that case, each of us will be doing the most good we can, given what the other person is doing. That is, given that you do A, I produce the best results by doing A, and vice versa. Thus, we will both succeed at doing what EA tells us to do. But together, we will be making things less good than we could have, since we could have made the outcome better if we had both done B.

The idea that EAs neglect institutions can be seen as an instance of this type of case. For example, suppose that EAs have to decide whether to send their charitable donations to GiveDirectly, which transfers money directly to extremely poor people, or to a political

[15] Allan Gibbard, 'Rule-Utilitarianism: Merely an Illusory Alternative?', *Australasian Journal of Philosophy* 43 (1965), pp. 211–20; Donald Regan, *Utilitarianism and Co-operation* (New York, 1980). The kind of example that Gibbard and Regan focus on is similar to the 'Hi-Lo' and 'Stag Hunt' examples discussed in economics and decision theory.

advocacy group lobbying to remove restrictions on immigration. For simplicity, let's again suppose that there are just two people, you and me. For each of us who donates to GiveDirectly, one person will ultimately be able to rise out of poverty. On the other hand, if we both donate to the advocacy group, the group will muster enough support to remove the immigration restrictions, which will have the effect of lifting millions of people out of poverty. But if only one of us donates to the advocacy group, this donation will accomplish nothing, and someone will remain in poverty whom we could otherwise have benefited. In this case, if we both donate to GiveDirectly, each of us will do the most good we can, given what the other is doing. Thus, we will succeed at doing what EA tells us to do. But again, we will together have failed to make the outcome as good as it could have been if we had acted differently.

It is important to note that this is not the only way to understand the institutional critique. As Berkey notes, for example, proponents of the institutional critique might be understood as arguing that EAs neglect important moral considerations that are not based on the idea of doing the most good we can, such as distinctive reasons of justice.[16] And while I have suggested simply that reforming institutions may be one way in which we can together do the most good, critics might think that the nature of certain institutions gives us special reasons to focus on them. For example, we might think that we have a collective obligation to reform our healthcare system, not simply because this is one way to do good, but because we have special obligations to provide for the basic needs of members of our own society.

My proposed interpretation of the critique, however, is one that EAs in particular have reason to find troubling. After all, many EAs have a broadly consequentialist outlook that is sceptical of distinctive reasons of justice, or, even if they accept such reasons, believe that they are not as strong as our reasons to alleviate the suffering of people in extreme poverty or of animals in factory farms. But it seems likely that EAs would not be happy if their efforts at addressing exactly these problems are not doing nearly as much good as they could be.

How do these cases motivate the appeal to collective obligations? When considering cases like this, as Derek Parfit writes, it seems 'in some sense obvious' what we ought to do: we ought to bring about the best outcome.[17] But we can't justify this claim by appealing to the idea that we individually ought to do the most good we can, because as we've seen, that will be true even if we bring about the second-best outcome.

[16] Berkey, 'Institutional Critique', pp. 163–4.
[17] Parfit, 'What We Together Do'.

A natural alternative is to understand the claim that we ought to bring about the best outcome as a claim about what we collectively ought to do. That is, we can understand this as a claim about an obligation possessed by the group as such, rather than merely a shorthand for claims about the obligations of the individual members of the group, such as obligations to try to cooperate with one another.[18] Thus, rather than focusing on the idea that each of us individually ought to do the most good we can, the critic can argue, EAs need to recognize that we collectively ought to do the most good we can.

Now, Holly Lawford-Smith has argued that it only makes sense to attribute obligations to groups that possess certain kinds of organization, and that are thus organized enough to be considered 'agents'.[19] This is because it is plausible to accept some version of the idea that 'ought implies can', and it is plausible that only agents can have the relevant kinds of ability. In contrast, other philosophers argue that even highly disorganized groups such as strangers witnessing an attack in a subway car, or humanity as a whole, can be capable of having obligations.[20] I will not attempt to resolve this dispute here. I will note, however, that even if we think that only highly organized groups are capable of having obligations, the various EA-aligned organizations could plausibly qualify. In addition, the EA community as a whole seems to have a significant degree of organization: EAs generally agree on a common set of values, and frequently work together in the various groups that are aligned with the movement, and in more general conferences. So on a somewhat looser standard, the EA community as a whole could also qualify as having obligations.

Are EAs actually facing a situation like those I have described? Do EAs actually have opportunities to collectively do more good by promoting institutional reforms that they are failing to pursue, because of their tendency to focus instead on giving to charities like GiveDirectly? This is a difficult empirical question. As of this writing, one prominent EA organization, Giving What We Can, reports that its members have so far donated about $25 million to charities they

[18] For a more detailed defence of this proposal, see Dietz, 'What We Together Ought to Do', pp. 960–3.

[19] Lawford-Smith, 'What "We"?', *Journal of Social Ontology* 1 (2015), pp. 225–49. See also Christian List and Philip Pettit, *Group Agency* (Oxford, 2011); Stephanie Collins, 'Collectives' Duties and Collectivization Duties', *Australasian Journal of Philosophy* 91 (2013), pp. 231–48.

[20] See Virginia Held, 'Can a Random Collection of Individuals Be Morally Responsible?', *The Journal of Philosophy* 67 (1970), pp. 471–81, at 476–7; Wringe, 'Global Obligations and the Agency Objection', *Ratio* 23 (2010): 217–31; David Killoren and Bekka Williams, 'Group Agency and Overdetermination', *Ethical Theory and Moral Practice* 16 (2013), pp. 295–307; Anne Schwenkenbecher, 'Joint Duties and Global Moral Obligations', *Ratio* 26 (2013), pp. 310–28.

believe to be highly effective. Critics might suggest that these EAs might instead have done more good by reserving this money to spend on political campaigns. For example, in 2016, they might have used this money to oppose the United Kingdom's referendum on leaving the European Union, and the presidential campaign of Donald Trump, both of which won by narrow margins, and both of which arguably have done and will continue to do serious damage on immigration, trade, and a variety of other institutional issues. On the other hand, it is hard to say whether even $25 million would have affected either of these outcomes. After all, Hillary Clinton's loss to Trump came despite the fact that her campaign outspent his by about $370 million.[21]

Even if EAs are not currently facing a situation like this, however, it is certainly possible that they will face such a situation at some point. After all, even if EA as a movement is not yet large enough to be able to have significant political influence, that could change. In addition, we have some reason to expect that even as EA grows, it may continue to largely avoid institutional causes. After all, one of EA's core recommendations is that individuals should focus on causes that are 'neglected', so that they can have a greater marginal impact.[22] But since efforts to affect institutions, such as political campaigns, often draw a lot of attention, this suggests that EAs might continue to avoid focusing on these efforts.

IV. WHY COLLECTIVE OBLIGATIONS NEED NOT AFFECT INDIVIDUAL OBLIGATIONS

Again, on my proposal, the problem for EA brought out by the issue of institutional reform is the idea that even if we succeed at doing what EA tells us to do, and individually do the most good we can, we may collectively fail to make the outcome as good as we could have made it. This problem shows why critics might plausibly argue that we need to recognize a collective obligation to do the most good we can. I will now argue that critics can make this appeal without running into the objections we saw earlier.

My proposal is simple. EA is deficient, critics can argue, in so far as it fails to recognize our collective obligations to do good. But we can do this without rejecting EA's core idea that each of us should do the most good we can. We just need to claim both that each of us should do the most good individually, and that we should do the most good

[21] Niv M. Sultan, 'Election 2016: Trump's free media helped keep cost down, but fewer donors provided more of the cash', *OpenSecrets.org*, <http://www.opensecrets.org/news/2017/04/election-2016-trump-fewer-donors-provided-more-of-the-cash/> (2017).

[22] MacAskill, *Doing Good Better*.

collectively. In other words, the problem is not that EA is mistaken, but rather that it is incomplete.

What Gibbard/Regan cases show is that even if we succeed at doing what EA tells us to do, we *could* fail to be collectively doing the most good. But these cases do not show that we have to choose between doing the most good individually and doing the most good collectively. After all, in our earlier example, if we both donate to immigration reform, then we each will have done the most good we can, given what the other is doing. And in fact, as Regan argues, whenever we produce the best outcome we can collectively, it must be true that each of us has produced the best outcome we could individually. This is because if I could have produced a better outcome by performing some alternative action, that would have to mean that we could have produced a better outcome by performing the set of actions that included this alternative.[23] So we can adopt the claim that we should collectively do the most good without rejecting the claim that we should individually do the most good.

This proposal can be understood as refocusing the institutional critique. Again, it is natural to understand the critique as claiming that individuals ought to try to work towards changes to the collective behaviour that we carry out through our institutions. In contrast, my proposal moves the focus from collective behaviour merely as an *object* of individual action to the idea of a collective *subject* of action: to thinking about the opportunities to do good that we have as a group.

This proposal also avoids the problems associated with both horns of our earlier dilemma. On this proposal, we can agree with EA that individuals should promote institutional reforms only if doing so would be the most effective thing they can do. We don't have to claim that individuals should promote institutional reforms even when their efforts wouldn't make a difference to bringing about these reforms, so that they could instead do more good elsewhere. But we've still identified a real shortcoming in the core EA philosophy. Even if EAs are right about what individuals should do, they need to think more about what they collectively should do.

V. IS THIS PROPOSAL RELEVANT IN PRACTICE?

The proposal I have offered invites a number of possible objections. I will close by discussing one challenge that may be particularly relevant to and worrisome for my proposal, and offering a brief sketch of how we might respond.

[23] Regan, *Utilitarianism and Co-operation*, pp. 54–5.

The challenge is this: how, exactly, is recognizing collective obligations supposed to make a practical difference for EAs? After all, collective actions can only come about through individual actions. So if EAs are still committed to following their individual obligations, and if recognizing collective obligations to do good doesn't tell us anything about our individual obligations, how is this recognition supposed to influence their behaviour?

For example, an EA might protest that in a Gibbard/Regan case, each of us might know that the other will do A. And in that case, if we retain the idea that each of us should simply do what does the most good individually, then we will still produce the suboptimal outcome. So how can my proposal explain how we could avoid situations like this?

Of course, we might say that even if recognizing collective obligations would not make a practical difference, it is still important to do this for theoretical reasons. But if it were true that recognizing these obligations would not make a practical difference, this would significantly detract from the force of the institutional critique of EA. After all, EAs are primarily concerned not with purely theoretical questions but rather with how to do the most good in practice.

I suggest that we instead respond to this challenge as follows. In addition to being motivated to doing the most good they can individually, EAs could engage in what economists and decision theorists call 'team reasoning'.[24] The central idea of team reasoning is that I might decide to perform some action not on the grounds that this action is itself seen as ideal in some way, but on the grounds that this action is my part in some group action that is seen as ideal in some way. Theorists argue that team reasoning can help agents to solve coordination problems, such as Gibbard/Regan cases. Thus, if EAs take some group action to be ideal in virtue of being one that they have a collective obligation to perform, and they are team reasoners, this could lead them to coordinate.

Now, if I decide to perform some action because it is my part in some group action that we have an obligation to perform, this might be because I have decided that I have an individual obligation to do my part. But if our collective obligations need not make a difference to our individual obligations, as I have suggested, then this won't work. Luckily, there is no reason to assume that I will decide to do my part only if and because I believe that I have an individual obligation to do

[24] Robert Sugden, 'Thinking as a Team: Towards an Explanation of Nonselfish Behavior', *Social Philosophy and Policy* 10 (1993), pp. 69–89; Michael Bacharach, 'Interactive Team Reasoning: A Contribution to the Theory of Co-operation', *Research in Economics* 53 (1999), pp. 117–47; Bacharach, *Beyond Individual Choice: Teams and Frames in Game Theory*, ed. Natalie Gold and Robert Sugden (Princeton and Oxford, 2006); Natalie Gold and Robert Sugden, 'Theories of Team Agency', *Rationality and Commitment*, ed. Fabienne Peter and Hans Bernhard Schmid (Oxford, 2007), pp. 280–312.

so. After all, it seems clear that we do not in general perform actions only when and because we believe we are obligated to perform them.

Finally, it is crucial to recognize that being a team reasoner is not necessarily in competition with being motivated to do the most good individually. For one thing, as team reasoning theorists have argued, team reasoners can be 'circumspect': their decision to do their part can be sensitive to whether they expect other members of the group to do their parts.[25] More generally, EAs' commitment to doing the most good they can individually could serve as what Peter Railton calls a 'counterfactual condition': although they might have other motivations, such as a motivation to do their part, it might also be true that they would never knowingly act in a way that was incompatible with doing the most good individually.[26] For example, once we recognize that we together ought to contribute to immigration reform, team reasoning could lead each of us to do our parts, and donate. But it might also be true that each of us would not have decided to donate if we believed that this would prevent us doing the most good we could do individually.

Now, again, this is only a brief sketch of how we might respond to this challenge. There are difficult theoretical questions about the nature of coordination, and I do not claim to have resolved these questions. Instead, what I hope to have done is to shift the debate. In order to resist the institutional critique as I have presented it, EAs will have to defend substantial views in moral psychology and the theory of collective action. They cannot simply dismiss the critique on the grounds that it either does not raise any real problems for their view, or has implausible implications for what individuals ought to do.

Interestingly, EAs themselves seem to be starting to accept the practical importance of thinking about doing good in collective terms, rather than only in individual terms. In his opening talk at the 2017 global EA conference in San Francisco, MacAskill argued that as the community has grown, it has become more important to move from asking 'How do *I* do the most good?' to asking 'How can *we* do the most good?'.[27] In fact, the conference's theme was 'Doing Good Together'.[28]

alex.dietz@gmail.com

[25] Bacharach, 'Interactive Team Reasoning'.

[26] Railton, 'Alienation, Consequentialism, and the Demands of Morality', *Philosophy and Public Affairs* 13 (1984), pp. 134–71.

[27] MacAskill, 'Opening Talk', <http://www.eaglobal.org/talks/opening-talk-sf/> (2017).

[28] For helpful feedback and discussion, I would like to thank Amy Berg, Mark Schroeder, Jonathan Quong, and an audience at the Second Annual PPE Society Meeting.